Network equilibration and first-principles liquid water

M. V. Fernández-Serra* and Emilio Artacho

Department of Earth Sciences, University of Cambridge, Downing street, Cambridge CB2 3EQ, UK

Motivated by the very low diffusivity recently found in ab initio simulations of liquid water, we have studied its dependence with temperature, system size, and duration of the simulations. We use ab initio molecular dynamics (AIMD), following the Born-Oppenheimer forces obtained from density-functional theory (DFT). The linear-scaling capability of our method allows the consideration of larger system sizes (up to 128 molecules in this study), even if the main emphasis of this work is in the time scale. We obtain diffusivities that are substantially lower than the experimental values, in agreement with recent findings using similar methods. A fairly good agreement with $D(T)$ experiments is obtained if the simulation temperature is scaled down by $\approx 20\%$. It is still an open question whether the deviation is due to the limited accuracy of present density functionals or to quantum fluctuations, but neither technical approximations (basis set, localization for linear scaling) nor the system size (down to 32 molecules) deteriorate the DFT description in an appreciable way. We find that the need for long equilibration times is consequence of the slow process of rearranging the H-bond network (at least 20 ps at AIMD's room temperature). The diffusivity is observed to be very directly linked to network imperfection. This link does not appear an artefact of the simulations, but a genuine property of liquid water.

I. INTRODUCTION

After ten years of success of DFT-based AIMD simulations of liquid water, including work on structural, dynamical, chemical and electronic properties,[1,2,3,4,5,6,7] recent papers[8,9,10] have questioned some of the results of those early studies, showing that if the simulations are allowed to run longer, the diffusivity drops by one order of magnitude and the liquid becomes over-structured. The origin of the discrepancy with experiments[11,12,13] is still unclear.

There are obvious limitations in the AIMD description of liquid water that could account for them, including the inability of present gradient-corrected (GGA) density functionals to describe dispersion interactions, or the complete neglect of quantum fluctuations in the classical treatment of nuclear dynamics. The former problem has hardly been addressed and demands simulations where van der Waals interactions are explicitly accounted for. The recent DFT proposals that include these interactions[14,15] are still too demanding to allow realistic AIMD simulations of this sort. Empirical force fields have an enormous advantage here, since those interactions can be reasonably well described with little effort. For the latter problem, the complete quantum mechanical treatment of both electronic and ionic degrees of freedom is still computationally too costly, and, even though some pioneering studies have recently appeared,[16,17] their approximations have to be pushed to the limit and their reliability is still unclear.[9] Empirical simulations including proton quantum effects are again much more feasible.

Even if a wide range of empirical potentials exist for pure liquid water that offer a better description than that attainable nowadays by DFT, it is extremely important to have a working description of liquid water at the DFT level, not for the study of water itself, but for that of systems interacting with water. This is important in scientific fields as wide as wet chemistry, biochemistry, geochemistry and environmental sciences. Empirical potentials do not have enough flexibility and transferability to describe the large variety of processes happening in wet systems.

The purpose of this work is to assess the situation regarding the equilibrium description of DFT water, as well as understand the equilibration process that lurks behind the problems in reaching it. We present in the following results of simulations for different sizes, at different temperatures, and for relatively long times. Because the long-term aim is using DFT water in interaction with other systems, the scalability of the DFT description becomes crucial. We have thus used a method based on numerical atomic orbitals of finite support that allows linear-scaling DFT calculations[18,19] and therefore the possibility of much more efficient treatment of larger system sizes. The method is validated below for DFT liquid water, including the localization approximations required for linear scaling.[19,20,21] After the characterization of the equilibrium properties, we present results on non-equilibrium relaxation processes, which provide insights into why the simulations need longer times, how to look at the DFT deficiencies, and, more importantly, into the nature of liquid water itself.

II. METHOD

Our simulations are performed using the self-consistent Kohn-Sham approach[22] to density-functional theory[23] in the generalized-gradient approximation (GGA). The BLYP[24,25] exchange-correlation functional was chosen following previous studies,[1] even if the reported performance for liquid water seems to be quite similar among GGA functionals[8] (a more detailed comparison using dif-

TABLE I: Selected properties of the water monomer and dimer calculated with our method for three different basis sets. They are compared to results of plane-wave (PW, 70 Ry energy cutoff) and well converged Gaussian (GTO) calculations, and to experiment. The BLYP functional was used in all the calculations. d stands for dipole moment, E_b for binding energy and BSSE for basis-set superposition error.

	Monomer			Dimer			
Basis	r_{OH}(Å)	$\widehat{HOH}$	d(D)	r_{OO}(Å)	$\widehat{OHO}$	E_b(eV)	BSSE
$DZP^{0.5}$	0.967	104.4	2.13	2.92	175.1	4.76	12%
$DZP^{0.2}$	0.970	104.2	2.04	2.94	177.1	4.68	20%
$DZP^{0.0}$	0.974	104.0	1.89	2.95	178.1	4.01	30%
PW^a	0.973	104.4	1.81	2.95	173.0	4.30	-
GTO^b	0.972	104.5	1.80	2.95	171.6	4.18	-
Exp.	0.957^c	104.5^c	1.85^d	2.98^e	174.0^e	5.44^f	-

[a]Ref. 1; [b]Ref. 9; [c]Ref. 31; [d]Ref. 32; [e]Ref. 33; [f]Ref. 34.

ferent functionals will be presented elsewhere).

Core electrons are replaced by BLYP-generated norm-conserving pseudopotentials[26,27] in their fully non-local representation.[28] Numerical atomic orbitals (NAO) of finite support are used as basis set, and the calculation of the self-consistent Hamiltonian and overlap matrices is done using the linear-scaling SIESTA method.[18,19] Integrals beyond two-body are performed in a discretized real-space grid, its fineness determined by an energy cutoff of 150 Ry .

The solution of the eigenvalue problem is performed either with the linear-scaling solver of Kim et al.[21] or by diagonalization. The former is more efficient for larger sizes (due to the cube scaling of the latter), but imposes an additional localization approximation on the basis functions for the occupied Hilbert space.[20,21] The effect of this approximation in our system is assessed below.

The NAO bases were variationally optimized for the water molecule.[29,30] The double-ζ polarized (DZP) level was found to offer a good balance between accuracy and efficiency. For this system it means two $2s$ orbitals, two $2p$ shells and one $3d$ shell for oxygen, and two $1s$ orbitals and one $2p$ shell for hydrogen. Three basis sets were tried at this level, differing in the cutoff radii of the support regions of their wave-functions. These radii are controlled by a single "pressure" parameter,[30] by which tighter orbitals are obtained with higher pressure. Three pressures were considered (0.0 GPa, 0.2 GPa, and 0.5 GPa), and their performance in the description of the water monomer and dimer is shown in Table I.

For the AIMD simulations we opted for the intermediate basis set (0.2 GPa), even if some numbers in the table are better reproduced by the long one (0.0 GPa). On one hand, the efficiency of the calculations rapidly degrades with longer basis-function cutoff radii. On the other, the importance of long basis-function tails for the monomer and dimer stems from their gas-phase character. In the liquid phase these long tails become irrele-
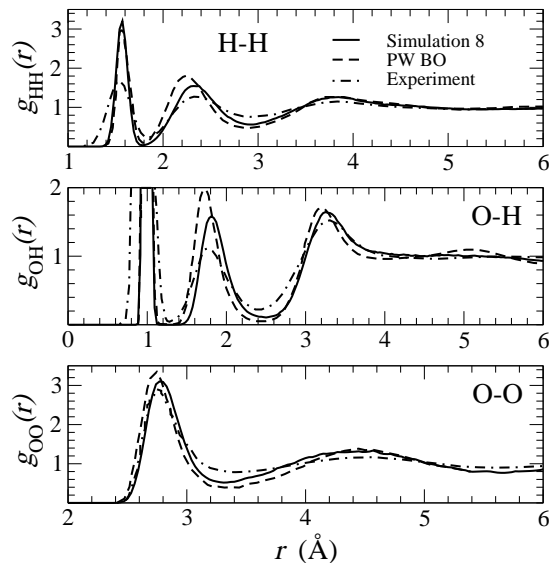

FIG. 1: Comparison of the H–H, O–H, and O–O radial distribution functions as obtained in this work for 64 molecules (solid line), with plane waves[9] (dashed line, PW BO stands for plane waves using Born-Oppenheimer forces), and by experiment[12,13] (dot-dashed line), at a temperature of 300 K.

vant given the presence of basis functions in neighboring molecules. This effect is quite apparent in the behavior of the dipole moment. Table I shows a clear tendency of growing overestimation of the dipole moment with tighter basis functions. However, the 2.04 D value obtained for the intermediate basis decreases to 1.74 D if calculated with the same basis set, but surrounded by the basis functions of four neighboring (absent) molecules. The long basis has the additional disadvantage of a large basis set superposition error (BSSE) in the hydrogen-bond (HB) description, while the short basis already shows appreciable discrepancies that are unlikely to be corrected by the presence of NAOs in neighboring molecules.

In Fig. 1, the results for radial distribution functions (RDFs) for liquid water as produced by our method, are compared with those of experiment[12,13] and those of recent AIMD Born-Oppenheimer results,[9] using a very similar method based on BLYP, norm-conserving pseudopotentials and a plane-wave (PW) basis. There are clear discrepancies between the two theoretical methods, our results showing an overall less structured liquid. Our slightly longer O-O distance along the HB seems to correlate with a shorter intramolecular O-H distance (a weaker HB is expected for a stronger intramolecular bond).

The differences have to be ascribed to incomplete basis sets, certainly in the NAO side, but also quite likely in the PW side. PW cutoffs in the range of 90 Ry or lower, as used in many PW studies (a 85 Ry cutoff was used in Ref. 9), are not extremely converged for GGA norm-conserving oxygen pseudopotentials, but rather represent a sensible compromise between accuracy and efficiency. Most important, however, is the fact that both simu-
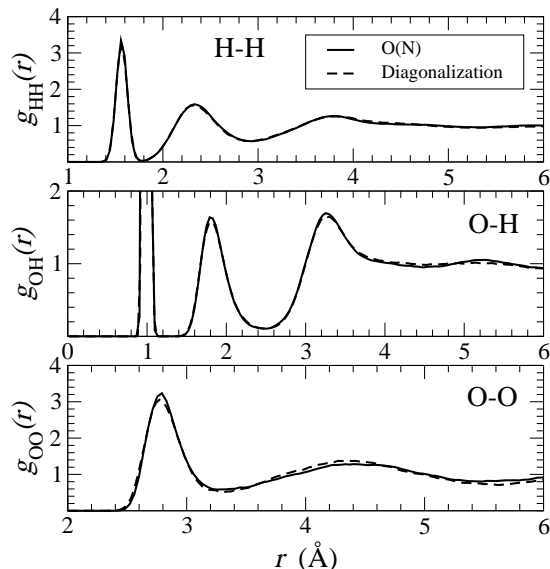
FIG. 2: Comparison of the RDFs obtained with diagonalization (solid line) and the $O(N)$ solver of Kim *et al.*[21] (dashed line) at a temperature of 300 K.

lations deviate from experiment in a very similar way, both displaying a clearly over-structured liquid. This over-structuring trend correlates with the very low diffusivities found by both methods, as discussed below.

The approximations above allow the linear-scaling computation of the Hamiltonian and overlap matrices. The $O(N)$ equivalent of the diagonalization demands an additional localization approximation. The $O(N)$ solver of Kim *et al.*[21] does it by imposing finite support to the Wannier-like localized solution wave-functions it finds.

Figure 2 tests the effect of the extra localization approximation by comparing the RDFs obtained using diagonalization and using the $O(N)$ solver in a 64-molecule 15-ps simulation. The confinement radius for the localized solution wave-functions is 5.0 Å. The $O(N)$ simulation corresponds to simulation 6 in Table II and the diagonalization one to simulation 8. The comparison is very satisfactory, the differences being substantially smaller than when comparing with experiment. From a practical point of view, the point at which the linear-scaling solver begins to be of advantage computationally, for this system and this basis, is around 32 molecules. Since the main emphasis here is on long time scales, the simulations presented below are up to 128 molecules only. For these sizes, diagonalization is still affordable and is the method chosen for the simulations below, in our aim to provide here as clean-cut results as possible. It is, however, an important consequence of this study, the new perspectives opened by the possibility of using linear-scaling for larger wet systems.

We have carried out a series of nine MD simulations of heavy water with varying size, density, temperature, and equilibration process. Their details are given in Table II. AIMD equilibration is accomplished by means

TABLE II: AIMD simulations performed in this work. N stands for the number of molecules, a for the box size, T for final equilibrated temperature, τ_{sim} for the AIMD simulation time after AIMD equilibration, τ_{eq} for the AIMD equilibration time, "Model" for the model used for preparation, T_{pre} for the temperature at which the preparation model had been equilibrated and T_i for the AIMD initial temperature (after the τ_{eq} anneal). Temperatures in K and times in ps.

#	N	a (Å)	T	τ_{sim}	τ_{eq}	Model	T_{pre}	T_i
1	32	9.865	298	20	4	AIMD	315	300
2	32	9.865	315	32	6	SPC/E	300	300
3	32	9.865	325	20	4	AIMD	300	335
4	32	9.865	345	30	4	TIP5P	325	325
5	32	9.890	305	20	4	AIMD	315	300
6	64	12.417	297	15	4	AIMD	320	300
7	64	12.417	326	24	4	SPC/E	315	300
8	64	12.460	303	20	4	AIMD	320	300
9	128	15.710	320	11	3	SPC/E	300	300

of temperature annealing (velocity re-scaling),[35] while the actual simulations are performed by straight Verlet's integration,[35] given our interest in dynamical quantities. In all the simulations (both empirical, see below, and ab initio) the time step used was 0.5 fs. The observed total-energy drifts corresponded to drifts in the system temperature between 0.26 K/ps and 0.36 K/ps. The different (final) temperatures in Table II are the result of different relaxation processes, not only in the preparation and further AIMD equilibration, but, most importantly, during τ_{sim} itself. Instead of the usual approach of long enough equilibration times and only monitoring the trajectories once well equilibrated, we chose to explore the long time-scale equilibration process itself, by monitoring the non-equilibrium part of the simulation, as discussed below.

The simulations were performed at constant volume (fixed cell size and shape, under periodic boundary conditions). The slight dispersion in the system densities considered in the literature suggested the study of the effect of slight density changes (below 1%) in the dynamical and structural results for our DFT water. Consistently with expectations (see Ref. 36 for the density dependence of the diffusivity), we found such effects of marginal importance for the present study and will not be further discussed here.

Empirical simulations were performed using different force fields (TIP5P[37] and SPC/E[38]) in order to prepare reasonably equilibrated starting points for AIMD. All these simulations were performed with the GROMACS MD package[39,40] under constant volume and temperature conditions using a Berendsen-type thermostat.[41] We equilibrated the simulations during 200 ps for 32 and 64 molecules and 150 ps for 128 molecules.
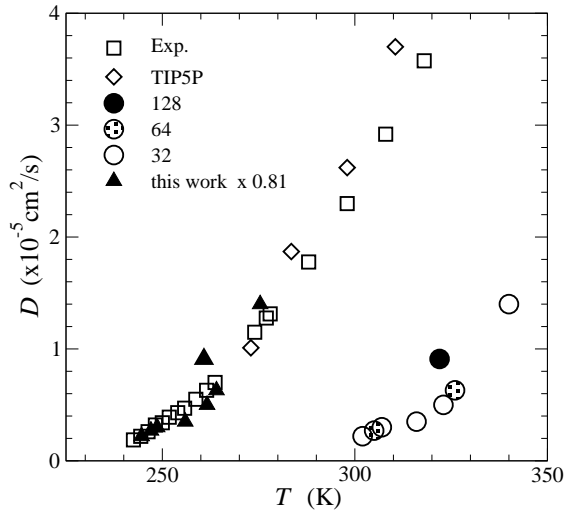
FIG. 3: Diffusivity vs temperature for this work (circles), experiment[11] (squares), TIP5P[42] (diamonds) and our AIMD data with a temperature re-scaling of 19% (triangles).

FIG. 4: Comparison of the H–H, O–H, and O–O RDFs as obtained in this work at 345 K (solid line), and by experiment[12,13] at 300 K (dashed line).

III. RESULTS AND DISCUSSION

A. Temperature dependence

In Figure 3 the temperature dependence of our computed diffusivity is presented and compared with experimental values at similar and lower temperatures[11] and the corresponding values for the TIP5P potential.[42] The diffusion coefficient is calculated using the Einstein relation:

$$6D = \lim_{t \to \infty} \frac{d}{dt} \langle |r_i(t) - r_i(0)|^2 \rangle. \tag{1}$$

Eq. (1) is evaluated computing the mean square displacement (MSD) of the oxygen atoms for multiple initial configurations equally spaced by 5 fs.

Confirming previous results,[8,9] the figure displays an underestimation of the room-temperature diffusivity of around one order of magnitude. It can also be seen as an overestimation of the temperature needed in the simulation to reach a certain diffusivity. Indeed, the same AIMD diffusivities plotted against a 19%-scaled-down simulation temperature give a quite acceptable agreement with experiment (triangles in the figure). The implication is that our AIMD description overestimates by around 25% the features of the energy landscape relevant for diffusion and flow.

One could then perform AIMD simulations at a higher temperature in order to describe a liquid with a diffusivity comparable to room-temperature experiments. In Fig. 4 we compare the room-temperature experimental RDFs with our corresponding results for a temperature of 345 K (the highest considered here), and the improvement is evident. It is unclear, however, what kind of agreement one would obtain for other properties, since
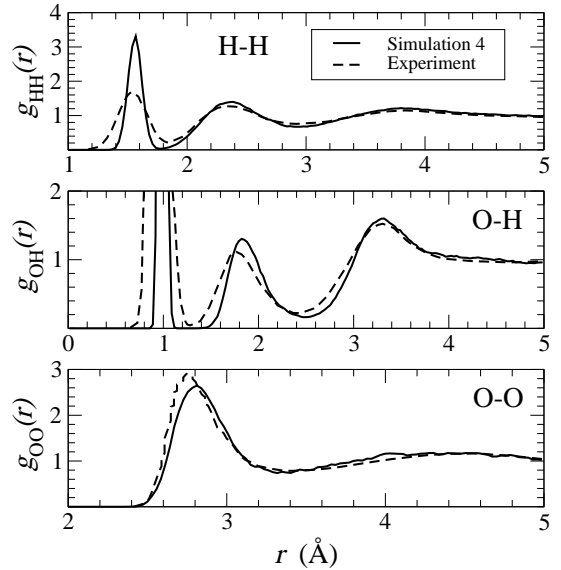
the local dynamics is controlled by atoms moving at velocities corresponding to the actual temperature of the simulations.

Schwegler et al.[9] reported a temperature overestimation that required scaling down by 28% (25% if the simulations use the Car-Parrinello scheme), slightly larger than our 19%, but clearly displaying the same trend. The higher diffusivity obtained with our NAOs as compared with PW, correlates with the less structured RDF for the NAOs in Fig. 1. Both discrepancies point to a weaker HB in the liquid for our description, as discussed in the previous section.

The most important result, however, is that both NAOs and PW disagree with experiment more substantially than among themselves. This shows that the main problem with the DFT description of liquid water is not in the technical approximations used in either method, but in the more fundamental approximations, namely, the GGAs (BLYP in this case) and/or the neglect of quantum fluctuations.

We experimented with different flavors of GGA, and in general we agree with previous reports in the conclusion that results for PBE[43] do not change the main findings for BLYP. It is clear that semi-local exchange-correlation functionals like these ones miss the non-local correlation effects that give rise to dispersion forces. However, since the origin of the discrepancy has not been ascribed to non-local effects, we decided to try other flavors, in the spirit of finding an efficient approach that works. A detailed study will be presented elsewhere, but preliminary results for the variant version of PBE proposed by Hammer et al.[44], called RPBE, showed promisingly higher diffusivities ($\approx 1.4 \times 10^{-5}$ cm^2/s at 300 K), with RDFs very similar to experiment if not under-structured.
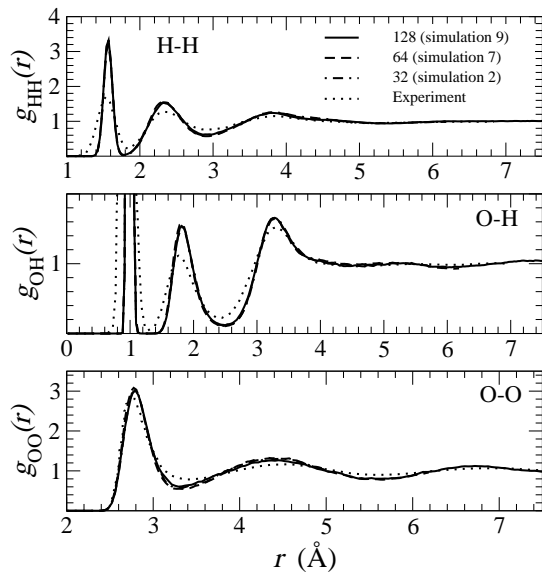
FIG. 5: Comparison of the H–H, O–H, and O–O RDFs obtained using 128 molecules (solid line), 64 (dashed line), 32 (dot-dashed line) and experiment[12,13] (dotted line).

B. Size effects

System-size effects have been addressed before[8,13] using mainly empirical potentials, due to the difficulty of studying larger sizes with AIMD. These studies indicated that size effects seemed not to be the problem, but the need for confirmation of this conclusion using AIMD was pointed out.[8,13] Not only because of the difference between empirical and ab-initio forces, but also because empirical MD approaches tend to impose radial cutoffs to the attractive R^{-6} potentials in order to avoid that an atom interacts with its periodic images. Simulations for the 32-molecule system impose quite a substantial and abrupt cut in the interactions, which originates a specific size effect, irrelevant to our problem, and possibly masking the interesting effects.

We have thus carried out simulations with 32, 64 and 128 water molecules. A comparison of their RDFs is shown in Fig. 5, where we do not find significant differences between the three sizes, supporting the conclusions of previous studies.[8] Furthermore, even if the 64-molecule and the 128-molecule simulations give slightly higher diffusivities than the 32-molecule one, it is apparent in Fig. 3 that the size effect produces marginal errors in the diffusivity as compared with the more substantial ones discussed before.

This absence of more substantial size effects could be taken as indication that the structure of water is mainly determined by short range forces.[8] The validity of such conclusion depends on what is understood by "structure". We have to keep in mind that in these simulations the density is fixed to the experimental value. An AIMD simulation of variable cell size (a large system size would
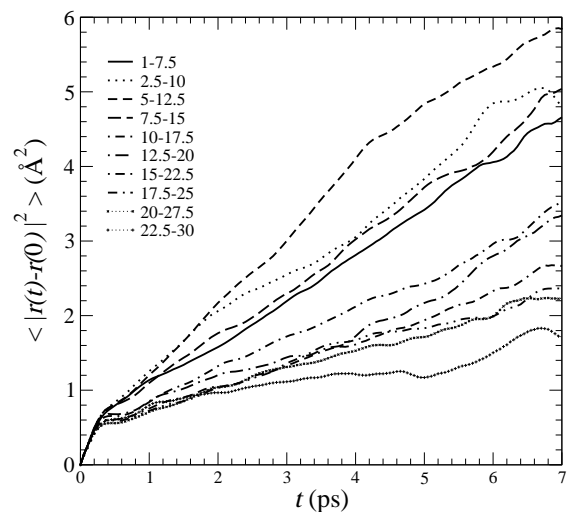

FIG. 6: Mean square deviation [as $F(t)$ in Eq. (2)] for the oxygen atoms vs time for simulation 2. The plots are computed in 7.5-ps-wide windows every 2.5 ps.

be required to reduce the pressure and volume fluctuations) would give a theoretical density that could appreciably differ from the experimental one, not least because of the neglect of dispersion forces. Indeed, we do observe in our simulations an average positive value of 2-4 kBar for the simulated pressure. Changes in the density will not show anything of substance in the nearest-neighbor peaks of the RDFs, since they are mainly determined by the HBs, but they will affect farther structural features.

C. Equilibration

It took ten years to find out about the problem of DFT water discussed above. That fact in itself points to a different problem, or rather a combination of two, both addressed in this section: (i) there is a long time scale associated to specific relaxation processes, and (ii) it can be difficult to observe them. Starting by the latter, it is customary to obtain the diffusivity D as in Eq. (1), from the slope of the function $F(t)$ defined in the form of an integral of the MSD instead of the MSD itself,

$$F(t) = \int_{t_r=0}^{t_{max}} \langle |r_i(t - t_r) - r_i(t_r)|^2 \rangle \qquad (2)$$

over a time window (t_{max}), normally the total of the simulation time. This procedure produces a substantial reduction of the fluctuations allowing a better determination of the required slope. If an equilibration process is still taking place and the diffusivity is intrinsically not stabilized yet, instead of a time evolution of that slope (a bend), still a (roughly) linear curve is obtained, with an average slope. It misleads to a reading of such slope as that of an equilibrated system.
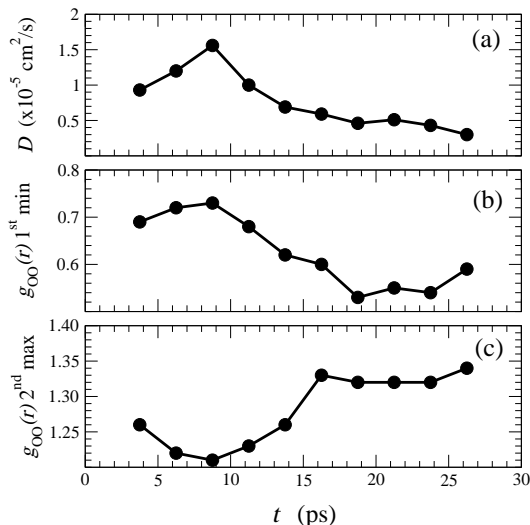
FIG. 7: Evolution of the diffusivity (a), the height of the first minimum (b), and of the second maximum for $g_{OO}(r)$ (c), in simulation 2. Calculated in 7.5-ps-windows every 2.5 ps.
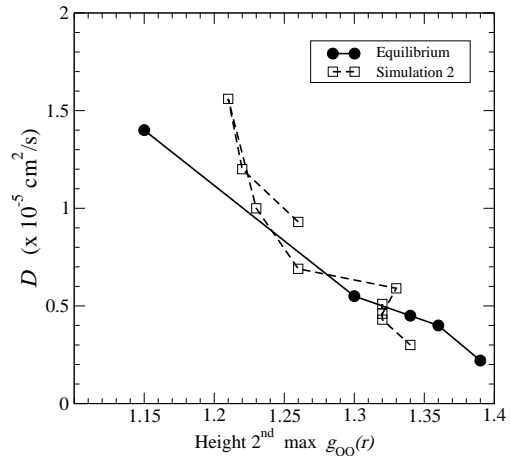

FIG. 8: Diffusivity versus height of the second maximum in $g_{OO}(r)$. Filled circles: equilibrium results for all the simulations with 32 molecules. Open squares: non-equilibrium evolution of the simulation 2 illustrated in Fig. 7.

Instead, we have chosen to consider our simulations in a sequence of overlapping time windows, each of them long enough (7.5 ps), to ensure enough range for the linear behavior to be extracted from the MSD averaged plot. The $F(t)$ curves are shown in Fig. 6. Obtaining a slope from them is still tricky and not devoid of ambiguities, including the choice of start and end times within the time window for the linear fit. We have checked, however, that, if systematic and carefully done, the variability in the extracted values does not affect the results in a substantial way for the purposes of this study. This "moving-window" approach allows observing the evolution of the diffusivity with time, and address the equilibration problem.

In Fig. 7 we show the evolution of the diffusivity and relate it to the evolution of the liquid structure, as monitored by he heights of the first minimum and of the second maximum of $g_{OO}(r)$. The figure shows a clear non-equilibrium behavior, both in the diffusivity and in the structural characteristics, in a time scale of tens of picoseconds, of the order 20-30 ps. Longer simulations would be needed to be more precise. We do not think it is justified at this point to determine the equilibration time more precisely, bearing in mind that it varies with temperature and possibly with size as well. A consequence of this result that should be kept in mind is that the "equilibrium" properties obtained in the previous section, as well as what obtained in previous studies (to our knowledge, the longest AIMD runs reported are not longer than 20-30 ps), are not necessarily completely equilibrated (as seen in Fig. 7), and should be taken with caution. History dependence is thus to be expected in similar AIMD simulations, i.e., dependence on the preparation model and initial temperature.

It is tempting at this stage to relate our equilibration

time with the one observed by inelastic UV scattering[45] for the structural relaxation probed by sound modes in the liquid, which, from values lower than 1 ps at room temperature, increases to higher than 20 ps below 250 K. Considering the 20% down-scaling discussed above, our room-temperature AIMD relaxation time scale would be consistent with the experimentally measured characteristic time for $\approx$240 K. If that is the case, accurate AIMD simulations (with no need for temperature re-scaling) should equilibrate in $\approx$1 ps. The same would be true for our AIMD at $T \approx$375 K.

Fig. 7 also shows a clear correlation between diffusivity and structure, very similar to what observed in equilibrium. This correlation is made explicit in Fig. 8, where the non-equilibrium behavior for the relaxation shown in Fig. 7 is compared with the equilibrium plot of D versus the height of the second maximum of $g_{OO}(r)$, as obtained in all the equilibrated simulations with 32 molecules.

This graph has interesting implications. The long equilibration time scale seems to be related to finding the equilibrium for some structural characteristic, X, but the diffusivity seems to equilibrate on a much shorter time scale to the instantaneous state of such X. This situation could be interpreted in terms of two time scales, one long, in which X evolves toward equilibrium, and a shorter one within which everything else happens, including diffusion.

Of course, this is an idealization, as is most clearly seen in the third point of Fig. 7(a), that corresponds to the highest open square in Fig. 8. This and the other smaller deviations from the equilibrium curve in Fig. 8 indicate that the equilibration of D to the corresponding value for $X(t)$ is not strictly instantaneous. This simple interpretation offers fruitful insights, however, which are explored in the following, by analysing the liquid structure in terms of a HB network.[46]
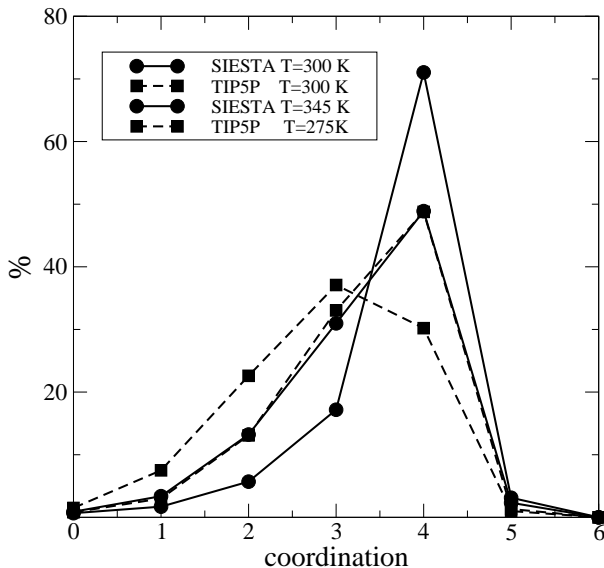
FIG. 9: Distribution of molecules with different coordinations. AIMD results (circles) are compared with results for the TIP5P potential[37] (squares). Comparison for the same temperature (300 K for both) is presented, as well as for 20% re-scaled temperatures (275 K for TIP5P and 345 K for AIMD; these are the ones practically superimposed).

D. Hydrogen-bond network

The molecules in water bind to each other by hydrogen bonds (HBs). Even if the character of this kind of bond is still controversial,[47,48,49,50] two important features are clear, namely, its strength (between that of a covalent bond and a Van der Waals one[46]), and its directionality. The chemical tendency is for each molecule to be surrounded by four others, donating two HBs and receiving two, in a tetrahedral arrangement. This tendency is perfectly satisfied in the crystalline phases of ice, but the HB network in liquid water at any given time is imperfect, with many four-coordinated molecules (even if the HBs may be stretched or bent), but some under-coordinated and over-coordinated ones as well. The picture is dynamic, with a continuous breaking and forming of HBs, with an average bond life-time of the order of 1 ps.[51] In this work we describe the HB network mainly by its coordination defects, namely, the under- and over-coordinated molecules.

In order to characterize the HB network a criterion must be adopted to decide whether two molecules are bonded or not. The usual criterion relies on two aspects, (i) an oxygen-oxygen distance smaller than some cut-off value, normally[52,53] the minimum after the first peak in $g_{OO}(r)$; and (ii) a HB angle larger than an arbitrary minimum value. Following Refs. 52,53 we adopted a minimum angle of 145°. For the characterization of network defects, we have also resorted to a temporal criterion for the definition of a HB, since fluctuations of distances or angles close to the critical values would otherwise appear
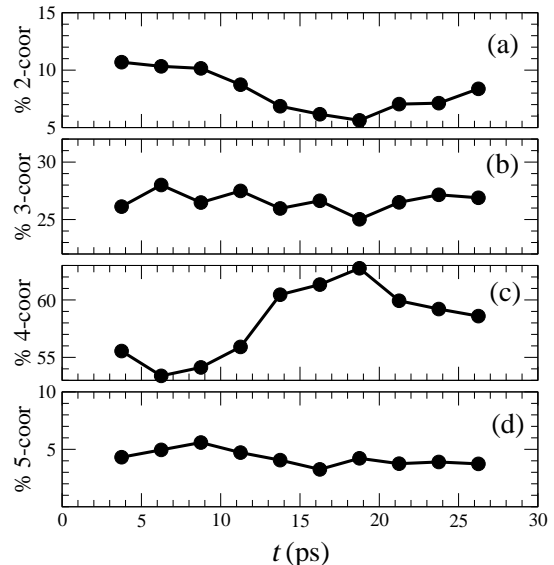

FIG. 10: Evolution of the proportion of molecules coordinated by two (a), three (b), four (c), and five (d) molecules. Averages done in 7.5-ps-windows every 2.5 ps (simulation 2).

as short-lived coordination defects, masking the defect statistics we want to monitor. Keeping track of HBs with life-times longer than typical vibrational or librational periods is enough for the purpose (we used a threshold of 250 fs, as defined by the cage effect[54]).

Fig. 9 shows the equilibrium distribution of molecules with different coordinations. Since there are no direct experimental results to compare to, comparison is presented with the results of the TIP5P force field. Note the asymmetry of the distribution, everything happening between the ideal coordination and under-coordination. Notice, however, that imposing a threshold life time in our HB definition biases the distribution toward lower coordinations (a longer minimum life time implies less HB qualifying as such, and thus the molecules are less coordinated). Nevertheless, we have checked the variation of the distributions with different life-time thresholds, finding that the overall shape is robust, maintaining the observed asymmetry. First, same temperatures are compared, both AIMD and TIP5P at 300 K, showing a very important difference, with AIMD displaying less than 30% defects while TIP5P shows 70%. Then, the AIMD at 345 K is compared with TIP5P at 275 K (-20%), where both distributions are, quite remarkably, hardly distinguishable.

Fig. 10 shows the evolution of the concentrations of molecules with different coordinations. For consistency we have used the same "moving window" approach as before. The figure shows that the relaxation process is clearly associated to changes in defect concentrations, and thus to a reorganization of the HB network. Furthermore, the non-equilibrium evolution of such defect concentrations (Fig. 10) is remarkably correlated with the evolution of the structural properties shown in Fig. 7.
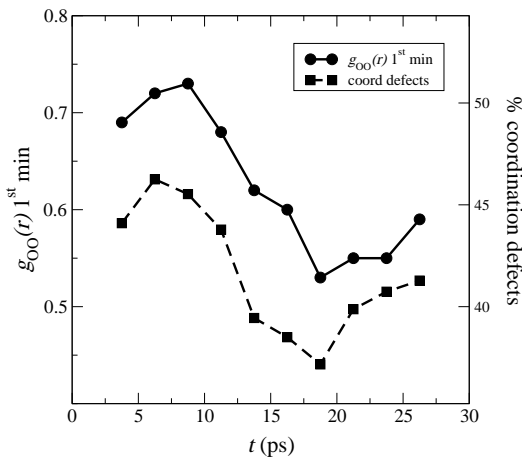
FIG. 11: Evolution of the first minimum of $g_{OO}(r)$ and the concentration of network defects, measured as % of molecules not tetra-coordinated (simulation 2).

This correlated behavior is further displayed in Fig. 11 where the dynamics of both properties are directly compared. Increases of four-coordinated molecules, with the consequent decrease in coordination defects, result in a reduction of the diffusivity and an enhancement of the structure in the RDFs. The under-coordinated molecules are the ones varying most, especially the bi-coordinated.

We have performed the same analysis for every simulation presented in this work. The general trend is as described, namely that increases in the concentration of under-coordinated (particularly bi-coordinated) network defects correlate with an increasing diffusivity of the system, confirming the link between diffusivity and network imperfection. The same link was found in Ref. 55, where the slow structural component of motion in supercooled water was associated to transitions between basins in the potential energy landscape. These transitions occur through changes in the local structure of the HB network.

Remarkably, the curve traced by a simulation in the diffusivity vs under-coordination plane is quite close and parallel to the equilibrium curve, obtained from joining the equilibrium points for the different simulations at different temperatures, as shown in Fig. 12. It is as if the diffusivity were mainly determined by the state of the network, the actual temperature becoming secondary. It is important to note as well that the state of the network is not completely characterized by just the concentration of coordination defects. The spatial distribution of such defects will also be relevant. This will be explored elswhere.

The evolution of the temperature along the simulations is consistent with the picture. When, for an initial temperature, the network is under-structured, the evolution toward more structuring (as in Fig. 7) is accompanied by increasing temperature, i.e., the network is finding regions of configuration space with lower potential energy. This is the situation when starting from empirical sim-
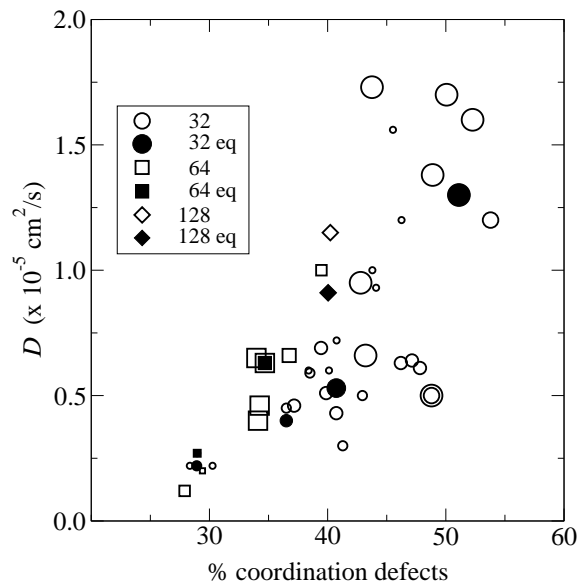

FIG. 12: Coordination defects versus diffusion coefficient (D) for all simulations. Open symbols represent non-equilibrium states as obtained in 7.5-ps-windows every 2.5 ps. Filled symbols represent equilibrium. Circles, squares and diamonds are for 32, 64, and 128 molecules, respectively. The size of each symbol reflects the (instantaneous) temperature, in five sizes, the smallest being for the range 295 K – 305 K, and the largest for 335 K – 345 K.

ulations equilibrated at the target AIMD temperature, since force fields tend to produce less structured liquids. If starting from an over-structured network, however, the slow increase in network defects is accompanied by a decrease in system temperature as the new defects are created. This is the case for simulation 3, for which the starting point was a previous AIMD simulation equilibrated at a lower temperature.

IV. CONCLUSIONS

We have carried out a series of AIMD simulations of liquid water using the SIESTA method, for temperatures between 300K and 350K. The conclusions of the present study can be summarized as follows:

(i) The differences between the SIESTA method at the level used in this study and PW-based AIMD methods[8,10] are less significant than the deviations between AIMD and experiment. This points to the more fundamental approximations (neglect of quantum fluctuations in the dynamics, and of dispersion forces) as responsible for the latter deviations.

(ii) The additional localization approximation imposed by the the linear-scaling solver produces errors much less significant than the ones mentioned above. This opens very good prospects for the study of complex systems in interaction with water.

(iii) The comparison of different system sizes (32, 64

and 128 molecules) shows marginal size effects for both RDF and diffusivity. The largest system seems to show a slightly higher diffusivity, to be confirmed by longer simulation times.

(*iv*) The AIMD results for RDF and diffusivity at a given temperature compare well with the experimental results for a temperature 20% lower, in fundamental agreement with PW AIMD results,[9] which require a lowering of around 28%. It means that the AIMD simulations performed in the past at room temperature were in fact describing supercooled water, with an effective temperature of around 240 K (at least for diffusivity and RDF purposes). The 20% temperature rescaling works also remarkably well for the comparison between AIMD and TIP5P in the HB network imperfection. Besides its fundamental implications, a direct practical consequence of this conclusion is that, if an AIMD simulation at temperature T is to be prepared by running an empirical model beforehand, it will be much more efficient to equilibrate the model to $0.8T$ rather than T.

(*v*) A slow equilibration process of a time scale of at least 20 ps at AIMD's room temperature has been identified. If it is related to the structural relaxation times characterized experimentally,[45] the process could have a much shorter time scale ($\approx$1 ps) at *real* room temperature, *i.e.*, for our AIMD simulations scaled up in temperature, or for AIMD with a better performing GGA.

(*vi*) In this equilibration process, the "instantaneous" diffusivity correlates with some instantaneous structural properties captured in the RDF, much more than with the actual instantaneous temperature.

(*vii*) HB network rearrangements have been proposed to be behind the slow equilibration process. Network imperfection (mainly the proportion of under-coordinated molecules) has been found to correlate very strongly with the diffusivity and the RDFs in their non-equilibrium evolution.

These last findings have important implications in the way we see the present DFT problems in the description of liquid water. Rather than being related to overestimation of energy barriers, this work points to an overestimation of the energy of formation of coordination defects. It is, therefore, not so much about transition states in the energy landscape as about the energy difference between basins in that landscape. Finally, we are convinced that the direct link described here between diffusivity and network imperfection is a property of liquid water, not an artefact of DFT.[55]

[*] Electronic address: mfer01@esc.cam.ac.uk
[1] M. Sprik, J. Hutter, and M. Parrinello, J. Chem. Phys. **105**, 1142 (1996).
[2] J. Ortega, J. P. Lewis, and O. Sankey, J. Chem. Phys. **106**, 3696 (1996).
[3] P. L. Silvestrelli and M. Parrinello, J. Chem. Phys. **111**, 3572 (1999).
[4] P. L. Silvestrelli and M. Parrinello, Phys. Rev. Lett. **82**, 3308 (1999).
[5] P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello, Science **291**, 2121 (2001).
[6] M. Boero, M. Parrinello, K. Terakura, T. Ikeshoji, and C. C. Liew, Phys. Rev. Lett. **90**, 226403 (2003).
[7] S. Izvekov and G. A. Voth, J. Chem. Phys. **116**, 10372 (2002).
[8] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, J. Chem. Phys. **120**, 300 (2004).
[9] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli, arXiv:cond-mat p. 0405561 (2004).
[10] D. Asthagiri, L. R. Pratt, and J. D. Kress, Phys. Rev. E **68**, 41505 (2003).
[11] R. Mills, J. Chem. Phys. **77**, 685 (1973).
[12] A. K. Soper, Chem. Phys. **258**, 121 (2000).
[13] J. M. Sorenson, G. Hura, R. M. Glaeser, and T. Head-Gordon, J. Chem. Phys. **113**, 9149 (2000).
[14] W. Kohn, Y. Meir, and D. E. Makarov, Phys. Rev. Lett. **80**, 4153 (1998).
[15] M. Dion, H. Rydberg, E. Schrder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).
[16] B. Chen, I. Ivanov, M. L. Klein, and M. Parrinello, Phys. Rev. Lett. **91**, 215503 (2003).
[17] S. Raugei and M. L. Klein, J. Am. Chem. Soc. **125**, 8992 (2003).
[18] P. Ordejón, E. Artacho, and J. M. Soler, Phys. Rev. B **53**, 10441 (1996).
[19] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, J Phys. Condens. Matter **14**, 2745 (2002).
[20] P. Ordejón, A. Drabold, M. P. Grumbach, and R. M. Martin, Phys. Rev. B **51**, 1456 (1995).
[21] J. Kim, F. Mauri, and G. Galli, Phys. Rev. B **52**, 1640 (1995).
[22] W. Kohn and L. J. Sham, Phys. Rev. **140**, a1133 (1965).
[23] P. Hohenberg and W. Kohn, Phys. Rev. **136**, b864 (1964).
[24] A. D. Becke, Phys. Rev. A **38**, 3098 (1998).
[25] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[26] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[27] The pseudopotential cutoff radii were $1.15a_0$ for all channels in O, and $1.20a_0$ in H. For O, a non-linear partial-core correction[56] was used in order to ensure a smooth

GGA pseudopotential near the nucleus.[57] It had the form proposed by M. M. G. Alemany and J. L. Martins (unpublished) $pc(r) = e^{\alpha + \beta r^2 + \gamma r^4}$, and a matching radius of $1.17 a_0$.

[28] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[29] J. Junquera, O. Paz, D. Sánchez-Portal, and E. Artacho, Phys. Rev. B **64**, 235111 (2001).

[30] E. Anglada, J. M. Soler, J. Junquera, and E. Artacho, Phys. Rev. B **66**, 205101 (2002).

[31] W. S. Benedict, N. Gailar, and E. K. Plyler, J. Chem. Phys. **24**, 1139 (1975).

[32] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman, J. Chem. Phys. **59**, 2254 (1973).

[33] J. A. Odutola and T. R. Dyke, J. Chem. Phys. **72**, 5062 (1980).

[34] L. A. Curtiss, D. J. Frurip, and M. Blander, J. Chem. Phys. **71**, 2703 (1979).

[35] M. P. Allen and D. J. Tildesley, *"Computer simulation of Liquids"* (Oxford University Press, 1987).

[36] I. Vaisman, L. Perea, and M. L. Berkowitz, J. Chem. Phys. **98**, 9859 (1993).

[37] M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).

[38] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Chem. Phys. **91**, 6269 (1987).

[39] Berendsen, H.J.C., van der Spoel, D., and R. van Drunen, Comp. Phys. Comm. **91**, 43 (1995).

[40] Lindahl, E., Hess, B., and D. van der Spoel, J. Mol. Mod. **7**, 306 (2001).

[41] H. J. C. Berendsen, J. P. M. Pstman, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[42] M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys. **114**, 363 (2001).

[43] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[44] B. Hammer, L. B. Hansen, and J. K. Norskov, Phys. Rev. B **59**, 7413 (1999).

[45] C. Masciovecchio, S. C. Santucci, A. Gessini, S. D. Fonzo, G. Ruocco, and F. Sette, Phys. Rev. Lett. **92**, 255507 (2004).

[46] F. Stillinger, Science **209**, 451 (1980).

[47] J.-H. Guo, Y. Luo, A. Augustsson, J. E. Rubensson, C. Sathe, H. Agren, H. siegbahn, and J. Nordgren, Phys. Rev. Lett. **89**, 137402 (2002).

[48] A. Romero, P. L. Silvestrelli, and M. Parrinello, J. Chem. Phys. **115**, 115 (2001).

[49] T. K. Ghanty, V. N. Staroverov, P. K. Koren, and E. R. Davidson, J. Am. Chem. Soc. **122**, 1210 (2000).

[50] M. V. Fernández-Serra and E. Artacho, to be published.

[51] A. Luzar, J. Chem. Phys. **113**, 10663 (2000).

[52] A. Luzar and D. Chandler, Nature **379**, 55 (1996).

[53] F. W. Starr, J. K. Nielsen, and E. Stanley, Phys. Rev. E **62**, 579 (2000).

[54] P. Gallo, F. Sciortino, P. Tartaglia, and S.-H. Chen, Phys. Rev. Lett. **76**, 2730 (1996).

[55] N. Giovambattista, F. W. Starr, F. Sciortino, S. V. Buldyrev, and H. E. Stanley, Phys. Rev. E **65**, 041502 (2002).

[56] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[57] D. R. Hamann, Phys. Rev. Lett. **76**, 660 (1996).